



Dual connectivity for LTE-advanced heterogeneous networks

Hua Wang¹ · Claudio Rosa² · Klaus I. Pedersen^{1,2}

Published online: 7 August 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Dual connectivity (DC) allows user equipments (UEs) to receive data simultaneously from different eNodeBs (eNBs) in order to boost the performance in a heterogeneous network with dedicated carrier deployment. Yet, how to efficiently operate with DC opens a number of research questions. In this paper we focus on the case where a macro and a small cell eNBs are inter-connected with traditional backhaul links characterized by certain latency, assuming independent radio resource management (RRM) functionalities residing in each eNB. In order to fully harvest the gain provided by DC, an efficient flow control of data between the involved macro and small cell eNBs is proposed. Moreover, guidelines for the main performance determining RRM algorithms such as UE cell association and packet scheduling are also presented. It is demonstrated how proper configuration of the proposed flow control algorithm offers efficient trade-offs between reducing the probability that one of the eNBs involved in the DC runs out of data and limiting the buffering time. Simulation results show that the performance of DC over traditional backhaul connections is close to that achievable with inter-site carrier aggregation (CA) and virtually zero-latency fronthaul connections, and in any case it is significantly higher compared to the case without DC.

Keywords Dual connectivity · Heterogeneous network · LTE-advanced · Radio resource management · Performance evaluation

1 Introduction

Heterogeneous networks (HetNets) consisting of macro and small cells are considered as one of the main steps towards meeting the future requirements for carrying the ever increasing broadband mobile traffic [1]. Although the migration to HetNet offers numerous benefits, it also introduces several paradigm shifts and challenges that call for new innovations and solutions to make it a true success [2, 3]. Recently, there are multiple research initiatives investigating further integration of macro and small cell functionalities to fully maximize the benefits of HetNet deployments. An overview of such techniques is provided in [4–6]. With different carrier frequencies deployed at macro and small cell layers, dual connectivity (DC) which extends the LTE-Advanced Carrier Aggregation (CA) functionality to allow user equipment (UE) to simultaneously receive data from both a macro and a small cell eNBs is a promising technique. DC is among the solutions standardized by 3GPP for Release 12 small cell enhancements. It aims to improve user throughput performance by utilizing radio resources in more than one eNB [7].

Due to the relatively short time since the introduction of DC in 3GPP Release 12, there is limited number of related studies in the open literature. The basic concept of DC is introduced in [8, 9]. The issues regarding the pairing of base stations and the grouping of mobile UEs are examined in [10]. From performance point of view, the user throughput and mobility benefits of DC in the form of inter-site CA have been analyzed in [11–14]. The energy efficiency with DC is evaluated in [15] in comparison with some of the existing traffic offloading mechanisms. However, those previous performance studies of DC [11–15] are mainly focused on the case where the small cells are realized with Remote Radio Heads (RRHs), assuming

✉ Hua Wang
huw@es.aau.dk

¹ Aalborg University, Aalborg, Denmark

² Nokia Networks, 9220 Aalborg Ø, Denmark

centralized base band processing at the macro, and virtually zero latency high-speed fiber-based fronthaul connections between macro and RRHs. In a more practical scenario, the macro and small cells are inter-connected via traditional backhaul connections (i.e., X2 interface in accordance with 3GPP LTE terminology) characterized by certain latency and limited capacity, assuming separate and independent radio resource management (RRM) functionalities [e.g. packet scheduling, medium access control (MAC), and Hybrid Automatic Repeat request (HARQ)] residing in each eNB. For such scenarios, data has to be forwarded from the macro cell eNB to the small cell eNB over the X2 interface before UEs configured with DC can benefit from data reception from the two cells. On the other hand, data transmission via the small cell will introduce additional delay due to the X2 latency and the buffering time in the small cell eNB. Therefore, efficient flow control management of data between the involved macro and small cell eNBs plays an important role.

In this paper, we focus on the case of DC over traditional backhaul. The main contribution is of three folds. We first focus on derivation of an effective inter-eNB flow control algorithm that aims at exploiting the full potential gain of DC while minimizing data buffering time in the small cell eNB. The proposed scheme keeps track of fast variations of the UE throughput and buffer status in the small cell eNB and works effectively under different backhaul configurations (e.g. X2 latency, flow control periodicity). Secondly, we provide guidelines for the design of performance-determining radio resource management (RRM) functionalities for DC such as UE cell association (i.e. how to configure UEs with DC) and packet scheduling (i.e. how to schedule the UEs configured with/without DC) in order to ensure the proper operation with DC. Thirdly, we present an extensive performance analysis under realistic conditions. In order to ensure high degree of realism and practical relevance of the results, the corresponding performance of the proposed algorithms should be evaluated under realistic multi-cell, multi-user conditions, using random point process deployment models of small cell nodes [16, 17], state-of-the-art stochastic radio propagation channel models, dynamic birth–death traffic models, and accurate representation of the many mechanisms that influence the performance. Due to the complexity of the system model and the various RRM elements involved, strictly analytical derivation of theoretical expressions becomes intractable. The performance is therefore assessed by means of advanced system level simulations. When feasible, the produced simulation results are validated against simpler theoretical findings and results from other sources in the open literature. Thus, the produced statistical reliable simulation results form a solid basis for drawing mature conclusions with high degree of realism.

The rest of the paper is organized as follows: Sect. 2 outlines the basic HetNet DC concept assumed in this study, as well as a simple analysis on the gain mechanisms with DC. Sections 3 and 4 present the proposals for flow control and RRM algorithms, respectively. Performance results and the underlying modeling framework for the simulations are presented in Sect. 5. Finally, concluding remarks are summarized in Sect. 6.

2 Concept description

2.1 System model

Let us consider a scenario composed of a set of macro cells (denoted as \mathcal{M}) and a set of small cells (denoted as \mathcal{S}) deployed at two non-overlapping carrier frequencies f_1 and f_2 , respectively. Due to the higher radio propagation loss at higher frequencies, it is assumed that $f_1 < f_2$ to ensure good wide area coverage for the macro layer. The baseline reference is according to the Rel'8 LTE specifications, where UEs are connected and served by a single eNB only, i.e. can only receive data from either a macro or a small cell eNB. For cases with DC, it is assumed that the existing CA functionality (see [18, 19] for additional background) is extended to support users receiving data simultaneously from two eNBs (in line with Rel'12 specifications). The assumption is that such users have their Primary Cell (PCell) configured on the best macro eNB, with the option of also having a Secondary Cell (SCell) configured on the small cell eNB when feasible. Figure 1 illustrates how different users are either in DC mode between a macro and a small cell eNB, or served by a single eNB only.

Figure 2 illustrates further details on the downlink data flow for users in DC mode between a macro and a small cell eNB. As pictured, it is assumed that the data-flow to the UE is as follows; user plane data from the Core Network (CN) is first transferred to the macro eNB [operating as the master eNB (MeNB)]. In the macro eNB the data flow is split, so some data are transmitted via the macro (PCell) to the UE, while other data are transferred over the X2 interface to the small cell eNB [operating as the secondary eNB (SeNB)], and transmitted to the UE via the corresponding cell (SCell). Though in theory the roles of master and secondary eNB do not depend on the eNB's power class and can vary among UEs, we assume that the MeNB is always a macro eNB while the SeNB is always a small cell eNB. The X2 interface imposes latencies from few milli-seconds to several tens of milli-seconds depending on the implementation. In alignment with 3GPP assumptions, the MeNB and SeNB are assumed to have independent medium access control (MAC) entities and physical layer processing [7]. This implies that the macro

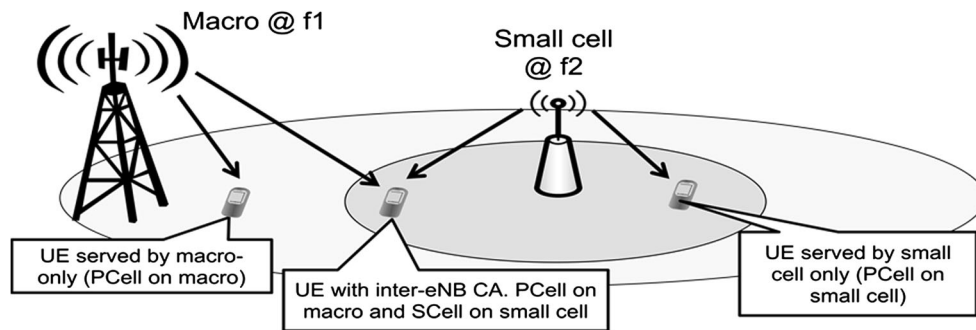


Fig. 1 Basic illustration of assumed scenario where UEs are either served by a single cell or are benefiting from DC by simultaneously receiving data from both a macro and a small cell eNB

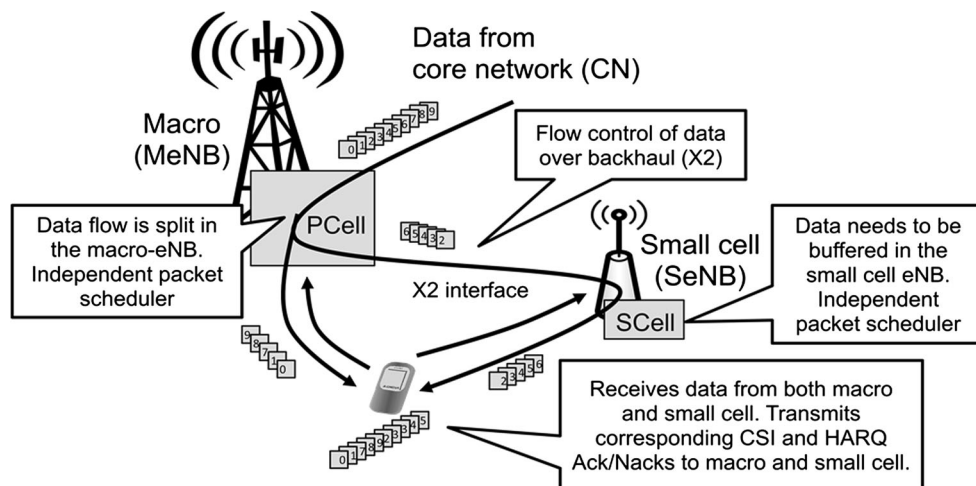


Fig. 2 High-level sketch of assumptions for a user in DC between a macro and a small cell

and the small cell eNBs each decide how to schedule data for the UE. Similarly, independent Hybrid Automatic Repeat request (HARQ) and link adaptation are assumed for the PCell and SCell transmissions in line with basic CA assumptions [18, 19]. The UE is assumed to have multi-carrier transmission capability for the uplink so that it can feedback separate Channel State Information (CSI) and HARQ (negative-) acknowledgements ((N)ACK) to the macro and small cell eNBs. Once the data packets have been decoded successfully by the UE, they are re-ordered and delivered to higher layers. It is therefore obvious that the performance of DC depends on multiple factors, where especially the design of RRM algorithms for deciding serving cell(s) for the UEs, packet scheduling, and flow control between the evolved eNBs over the X2 interface are of importance.

Notice that the LTE-A DC concept has some similarities with the multi-flow concept defined for High Speed Packet Access (HSPA), where users also can be served simultaneously by different base station sites [20, 21]. In the

further analysis and derivation of algorithms for LTE-A DC, we therefore strive towards exploiting findings from HSPA multi-flow studies when feasible.

2.2 Gain mechanisms with DC

The basic gain mechanism offered by DC is illustrated in the following for a simple single user case based on theoretical calculations. Let $i^m = \arg \max_{i \in M} \{R_i\}$ and $i^s = \arg \max_{i \in S} \{R_i\}$ denote the cell which offers the best estimated user throughput in the macro layer and small cell layer, respectively. R_i is the estimated user throughput of cell i using Shannon's capacity formula: $R_i = B_i \log_2(1 + \gamma_i)$, where B_i and γ_i are the available bandwidth and the Signal-to-Interference-plus-Noise Ratio (SINR) of cell i , respectively. For cases without DC, the user is assumed to be served by a single cell characterized by the best estimated throughput. The Shannon capacity for the user without DC can be expressed as:

$$C_{\text{noDC}} = \max_{i \in \{i^m, i^s\}} \{R_i\} \quad (1)$$

For cases with DC, the user is assumed to be served by both a macro cell and a small cell. The candidate cells characterized by the best estimated throughput in both the macro and small cell layers are selected as the serving cells. The Shannon capacity for the user with DC is expressed as:

$$C_{\text{DC}} = R_{i^m} + R_{i^s} \quad (2)$$

The user throughput gain with DC is:

$$G = \frac{C_{\text{DC}} - C_{\text{noDC}}}{C_{\text{noDC}}} \times 100\% = \frac{\min_{i \in \{i^m, i^s\}} \{R_i\}}{\max_{i \in \{i^m, i^s\}} \{R_i\}} = \frac{B_{i^q} \log_2(1 + \gamma_{i^q})}{B_{i^p} \log_2(1 + \gamma_{i^p})} \times 100\% \quad (3)$$

where $i^p = \arg \max_{i \in \{i^m, i^s\}} \{R_i\}$, and $i^q = \arg \min_{i \in \{i^m, i^s\}} \{R_i\}$. It is observed in (3) that the throughput gain depends on the channel quality (SINR) and the available bandwidths in the two layers. Let $\delta = \text{dB}(\frac{\gamma_{i^p}}{\gamma_{i^q}})$ and $k = \frac{B_{i^p}}{B_{i^q}}$ denote the SINR difference and the bandwidth difference between the two layers, respectively. Figure 3 shows the user throughput gain with DC for a single user case with different values of δ , k and γ_{i^p} . If the same bandwidth is deployed at the two layers (i.e., $k = 1$), it indicates that DC is most beneficial for users experiencing similar channel conditions in both layers (i.e., 100 % DC gain when $\delta = 0$). Notice the DC gain obtained from (3) cannot be larger than 100 %, due to the reason that for cases without DC the serving cell is selected with the highest estimated throughput from the candidate cells from the two layers.

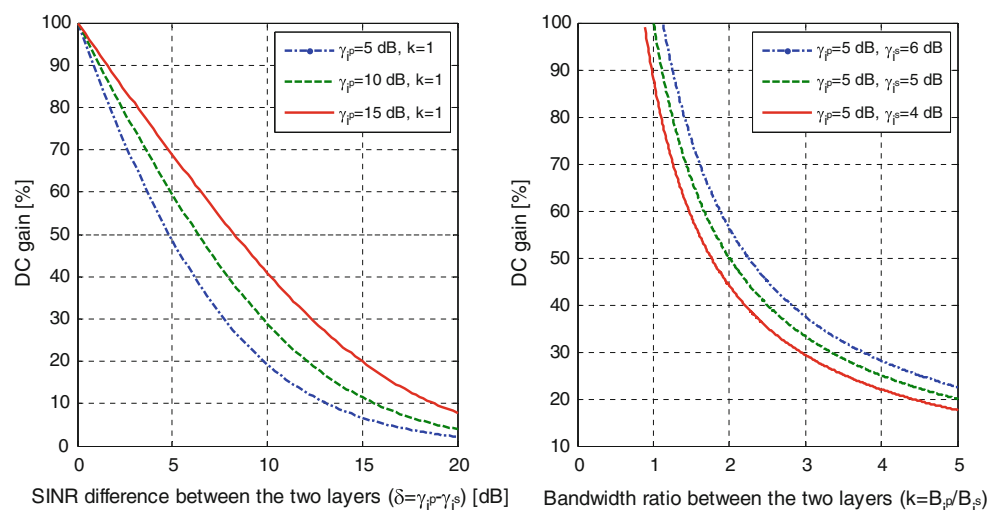
Although Fig. 3 offers useful insights of which users could potentially benefit from DC, it still remains to be further analyzed how DC performs in a more realistic

setting with a higher number of cells, varying number of users, etc. Thus in the following sections, we focus on the derivation of flow control algorithm between the involved macro and small cell eNBs for DC, as well as the performance-determining RRM functionalities.

3 Flow control

As shown in Fig. 2, the MeNB has to forward the data to the SeNB for UEs operating in DC mode. Received data from the MeNB are buffered in the SeNB until they are transmitted over the air interface to the UE via the SCell. Thus, the key question is how much data the MeNB should forward to the SeNB. If the MeNB doesn't forward enough data to the SeNB, the SeNB buffer may often run out of data, thus limiting the user throughput gain provided by DC. On the other hand, if too much data is pushed to the SeNB, buffering delay at SeNB is increased and it may even happen that the SeNB experiences buffer overflow while the MeNB buffer runs empty. The design target of the flow control algorithm is therefore to guarantee that there is always data to be scheduled in the SeNB so that UEs configured with DC can benefit from simultaneous data reception from the two cells, while limiting the probability of buffer overflow and reducing the additional delay introduced by transmission via the SeNB. Flow control can be implemented by different means such as window strategies and rate control schemes [22]. The target of flow control in DC is to match the data rate experienced in the SeNB. As the SeNB has the information of the scheduled user throughput of its associated UEs (both instantaneous and average value) as well as buffer status, it is a natural choice for the SeNB to decide how much data the MeNB should forward. Therefore in this paper, the flow control algorithm for data forwarding from the MeNB to

Fig. 3 DC gain with different values of δ , k and γ_{i^p} , single-user case



When user i is configured with DC, the MeNB forwards an initial amount of data to SeNB s . In the initialization phase of DC, the MeNB only has limited information such as UE measurement reports (e.g. channel quality

$$I_{i,s} = \min \left\{ \mathbf{E}_H [\log_2 \det (\mathbf{I}_{M_r} + \Gamma_{i,s} \mathbf{H} \mathbf{H}^H)] \cdot \frac{1}{N_s(t) + 1} \cdot B_s \cdot 2\Delta, I_{\max} \right\} \quad (4)$$

where I_{\max} is the maximum amount of data to be initially transferred to SeNB (I_{\max} is introduced to avoid too much data is transferred to SeNB in the initialization phase), B_s is the available bandwidth at SeNB s , $N_s(t)$ is the number of

active users at time instant t at SeNB s , and $\Gamma_{i,s}$ is the estimated wideband Signal-to-Interference-plus-Noise Ratio (SINR) of the user i at SeNB s , assuming that all eNBs are transmitting, calculated as:

$$\Gamma_{i,s} = \frac{P_s/M_t \cdot g_{i,s}}{\sum_{n \in S \setminus \{s\}} P_n \cdot g_{i,n} + N_o} \quad (5)$$

where P_s is the transmission power of SeNB s , $g_{i,s}$ is the downlink path gain from user i to SeNB s , and N_o is the background thermal noise. Though not completely true due to different users experiencing different relative radio conditions in macro cell layer as compared to small cell layers, for the sake of simplicity in (4), it is assumed that the resources are equally shared among the users which are schedulable for transmission in each cell.

At time instant t , the SeNB sends the data forward request to the MeNB. The data request is based on the target of maintaining the amount of data in the SeNB buffer to a level that can be transmitted in a predefined time interval. This time interval is set to be a configurable parameter denoted as θ_s , also referred to as target buffering time at the SeNB. In our proposed algorithm, the target is that the expected amount of data in the SeNB buffer at time instant $t + 2\Delta$, denoted as $\tilde{L}_{i,s}(t + 2\Delta)$, is equal to $\bar{R}_{i,s}(t) \cdot \theta_s$, expressed as follows:

$$\begin{aligned} \tilde{L}_{i,s}(t + 2\Delta) &= L_{i,s}(t) + K_{i,s}(t) + D_{i,s}(t) - \bar{R}_{i,s}(t) \cdot 2\Delta \\ &= \bar{R}_{i,s}(t) \cdot \theta_s \end{aligned} \quad (6)$$

where $L_{i,s}(t)$ is the actual amount of data stored in the buffer of user i at time instant t at SeNB s , $\bar{R}_{i,s}(t)$ is the average past scheduled throughput of user i at time instant t at SeNB s , $D_{i,s}(t)$ is the amount of data requested by SeNB s of user i at time instant t , and $K_{i,s}(t)$ is the amount of pending data forward requests till time instant t of user i at SeNB s .

The setting of θ_s determines the amount of data to be requested by the SeNB. For time invariant $\bar{R}_{i,s}(t)$ (e.g., constant channel condition and static resource allocation over time), the optimal setting of θ_s would be zero so as to minimize the buffering time in SeNB. However for time varying $\bar{R}_{i,s}(t)$ (e.g., the channel is subject to fast fading and with dynamic resource allocation), it is better to keep some extra amount of data in the SeNB buffer (i.e., $\theta_s > 0$) to compensate for the fast variations of instantaneous user throughput. The objective is to ensure that there is data buffered when the user is scheduled in the SeNB. By rearranging (6), the amount of data to be requested by SeNB s at time instant t for user i can be expressed as:

$$D_{i,s}(t) = \max\{0, \bar{R}_{i,s}(t) \cdot (2\Delta + \theta_s) - L_{i,s}(t) - K_{i,s}(t)\} \quad (7)$$

At time instant $t + \Delta$, the MeNB receives the forward data request transmitted by the SeNB at time instant t . The

MeNB forwards data to the SeNB only if the buffer size in the MeNB is larger than a certain threshold. The idea of setting a minimum buffer size in the MeNB is to prevent the MeNB from forwarding data to the SeNB if the MeNB can finish the data transmission faster than transferring the data to the UE via the SeNB. The minimum buffer size $T_{i,m}(t)$ at time instant t for user i at MeNB m is calculated as:

$$T_{i,m}(t) = \bar{R}_{i,m}(t) \cdot (\Delta + \theta_s) \quad (8)$$

where $\bar{R}_{i,m}(t)$ is the past average scheduled throughput to user i at time instant t at MeNB m , and $\Delta + \theta_s$ is the estimated time by transferring the data via SeNB. It is assumed that the SeNB and the MeNB keep records of the past average scheduled throughput per UE at the corresponding eNB for flow control purposes. If the remaining data in the MeNB is larger than $T_{i,m}(t)$, the MeNB forwards the requested amount of data to the SeNB. Otherwise, no data is forwarded to the SeNB.

At time instant $t + 2\Delta$, the SeNB receives the requested data from MeNB. The request-and-forward based flow control mechanism repeats periodically until the MeNB stops forwarding data to the SeNB either due to the completion of data transmission or because the buffer size in MeNB is below the threshold value in (8).

4 Radio resource management considerations

The two main RRM functionalities that determine the radio resource allocation among the users in the system are the cell association criteria and the packet scheduling. Hence, these RRM decisions are also the ones that impact the most on the performance (and relative gains) of applying DC. The assumptions for these two sets of RRM algorithms are therefore outlined and motivated in the following.

4.1 Cell association

The serving cell for a UE is determined based on downlink UE measurements. The UE measures the Reference Signal Received Power (RSRP) from each cell, as well as the Received Signal Strength Indicator (RSSI) on each component carrier. The RSRP expresses the received power of the transmitted reference signal from the different cells, while the RSSI is equivalent to wideband received power per carrier [23]. Expressed in decibel, the Reference Signal Received Quality (RSRQ) for one cell equals the RSRP minus the RSSI on the corresponding carrier. The UE can be configured to perform measurements of RSRP/RSRQ from its serving and surrounding cells. In dedicated carrier deployment, the RSRQ-based cell selection is preferred as it captures the channel quality and load conditions experienced on the corresponding layer [11]. For UEs not

configured with DC, the default assumption is that the serving cell for a user is selected as:

$$n^* = \arg \max_{n \in \mathcal{C}} \{ \text{RSRQ}_n + \text{RE}_n \} \quad (9)$$

where $\mathcal{C} = \mathcal{M} \cup \mathcal{S}$ is the set of candidate serving cells, and RE_n is the Range Extension (RE) for cell n , assuming $\text{RE}_n = 0$ for $n \in \mathcal{M}$ (macro cells) and $\text{RE}_n \geq 0$ for $n \in \mathcal{S}$ (small cells). Thus, adjusting the value of the RE for the small cells enables a simple form of inter-layer load balancing between the two frequency layers, offloading more UEs from the macro cell to the small cells [24]. In 3GPP terms, the criterion in (9) is roughly equivalent to using event A3 for user mobility (i.e. neighbor becomes offset better than PCell) [25]. An alternative to (9) is to apply a more opportunistic approach, where the UE simply connects to the small cell layer when the received RSRQ from a small cell is above a certain threshold (TH), and otherwise connects to the macro layer, i.e.,

$$n^* = \begin{cases} \arg \max_{n \in \mathcal{S}} \{ \text{RSRQ}_n \} & \text{if } \exists s \in \mathcal{S} : \text{RSRQ}_s > \text{TH} \\ \arg \max_{n \in \mathcal{M}} \{ \text{RSRQ}_n \} & \text{otherwise} \end{cases} \quad (10)$$

In 3GPP terminology, this opportunistic approach for connecting to the small cell is roughly equivalent to using event A4 for user mobility (i.e. neighbor becomes better than threshold) [25]. The rationale behind this approach is to have users offloaded to the small cell layer as soon as the quality on that layer is sufficiently good.

UEs configured with DC can be simultaneously connected to a macro and a small cell eNB. It is assumed that the UE has its PCell configured on the best macro cell (according to the specific cell association criteria), with the option of also having a small cell configured as SCell when feasible. The serving PCell is selected corresponding to highest received RSRQ from the macro cells, while the SCell is configured once the received RSRQ from the candidate cell (e.g., the cell with the highest received RSRQ from the small cells) is above a certain threshold:

$$\begin{aligned} n_p^* &= \arg \max_{n \in \mathcal{M}} \{ \text{RSRQ}_n \} \\ n_s^* &= \begin{cases} \arg \max_{n \in \mathcal{S}} \{ \text{RSRQ}_n \} & \text{if } \exists s \in \mathcal{S} : \text{RSRQ}_s > \text{TH} \\ \emptyset & \text{otherwise} \end{cases} \end{aligned} \quad (11)$$

where n_p^* and n_s^* are the selected PCell and SCell, respectively. Further studies on cell association and mobility management for cases with DC can be found in [13, 14].

4.2 Packet scheduling

The well-known Proportional Fair (PF) scheduler is applied, which is designed to converge to allocating equal

resource share to the users on average, assuming that the users exhibit similar fading variability [26]. The problem with baseline PF scheduler is that in scenarios where DC and non-DC UEs coexist, it will allocate more resources to the users connected to multiple cells (with DC) than users with single connectivity (without DC). For the example where user 1 is connected to cell A, user 2 is connected to cell B, and user 3 is connected to both cell A and B by using DC, the baseline PF scheduler will result in the following resource allocation: In cell A, users 1 and 3 will on average be allocated 50 % of the transmission resources. In cell B, users 2 and 3 will on average be allocated 50 % of the transmission resources. This results in an unfair resource allocation among the three users as user 3 is allocated twice the amount of resources as compared to users 1 and 2. As studied in [27, 28], the solution to this imbalance is to use a modified form of PF when calculating the scheduling metric. The scheduled user i^* on physical resource blocks (PRBs) j in cell k in subframe t is determined as:

$$i^*(t) = \arg \max_{i \in U(t)} \left\{ \frac{r_{i,j,k}(t)}{\sum_k \bar{R}_{i,k}(t)} \right\} \quad (12)$$

where $U(t)$ is the set of candidate users for scheduling in subframe t , $r_{i,j,k}(t)$ is the currently supported data rate for user i on PRB j in cell k (i.e. obtained from the CSI feedback), while $\bar{R}_{i,k}(t)$ is the past average scheduled throughput of user i from cell k in subframe t . The value of $\bar{R}_{i,k}(t)$ is estimated based on the past scheduled throughput for each user by using a first order autoregressive filter. The denominator of the PF scheduling metric for user i with DC equals the aggregated past average throughput of user i from all its configured cells (i.e. the user's total throughput). By applying this modification, it was shown in [27, 28] that the underlying utility function $\sum_i \log(R_i)$ is maximized also for the cases where some users are served only by one cell, while other users are served by multiple cells using CA functionality, resulting in more fair resource sharing among the users. In HetNet scenarios where the users can experience very different channel and load conditions on different carriers this modification is especially important as it tends to schedule the user on the cell with better channel conditions, thus improving the overall resource utilization efficiency. This type of scheduler is called cross-carrier PF as the scheduling metric for users in one cell also depends on the past average scheduled UE throughput on other cells. Hence, it is assumed that the packet schedulers in the macro and small cell eNBs periodically exchange information about the past average scheduled throughput for the users that are configured with DC between those cells. As the average scheduled user throughput is estimated over a relative longer time-window

(e.g. 400–500 ms) and hence does not vary on a fast basis, the rate of this information exchange among the involved eNBs can be on a moderate time-scale (e.g. every 50–100 ms) and is not sensitive to the transmission delay over X2-type backhaul connections.

5 Performance analysis

5.1 Simulation assumptions

The simulated environment is in line with 3GPP Rel-12 Scenario 2A as defined in [29]. The network topology consists of a standard hexagonal grid of three-sector macro eNBs complemented with a set of outdoor small cells. Macros and small cells are deployed at 2 and 3.5 GHz, respectively, assuming 10 MHz carrier bandwidth at each layer. A directional 3D antenna pattern with down-tilt is modeled for the macro cells, while small cells are simply equipped with omni-directional antennas. The transmission power for the macro and small cells is 46 and 30 dBm, respectively. The macro inter-site distance is 500 m. The small cells are randomly deployed in condensed clusters with 4 cells within a circular area with 50 m radius according to a uniform point process, subject to a minimum distance of 20 m between small cells. There is one small cell cluster per macro cell area. The ITU defined geometrical channel model is applied, where macro to UE links follow the Urban Macro model (UMa), while small cell to UE links are based on the Urban Micro model (UMi) [30]. Both UMa and UMi include separate models for line-of-sight (LOS) and non-LOS (NLOS). Selection between the LOS and NLOS model is random for each link, where the probabilities for selecting LOS or NLOS vary with the distance between the UE and eNB. Furthermore, notice that the UMa and UMi models are fairly advanced in the sense that effects such as shadow fading, angular dispersion, and temporal dispersion are correlated as also observed in [31] from analysis of field measurements.

The simulator follows the LTE specifications, including detailed modeling of major RRM functionalities such as packet scheduling, hybrid ARQ and link adaptation [32]. Closed loop 2×2 single user MIMO with pre-coding and rank adaptation is assumed for each link and the UE receiver type is Interference Rejection Combining (IRC) [33]. A dynamic birth–death traffic model is applied for generating user calls, where call arrival is according to a Poisson process with arrival rate λ per macro cell area. The hotspot deployment model is assumed, where 2/3 of the calls are generated in the small cell clusters, while the remaining UEs are generated uniformly over the entire simulation area. Each call has a finite payload size of $B = 4$ Mbits. Once the payload has been successfully

received by the UE, the call is terminated and the UE is removed from the simulation. Thus, the average offered load per macro cell area equals $\lambda \times B$. Channel-aware cross-carrier PF scheduler is used as specified in Sect. 4.2. The link to system mapping is based on the exponential effective metric model [34]. The non-ideal backhaul connections are explicitly modeled by assuming an X2 latency ranging from 5 to 50 ms. Flow control between the MeNB and SeNB is performed with the periodicity ranging from 5 to 20 ms. The schedulers in the macro and small cell exchange information on the past average scheduled throughput at the respective eNB for UEs that are configured with DC every 50 ms. Cases with inter-site CA and zero latency fronthaul connections (i.e. ideal flow control) are simulated as well in order to provide an upper performance bound for DC. Table 1 summarizes the main parameters used in the system-level simulations.

The main Key Performance Indicators (KPIs) are the 5 and 50 % downlink user throughput. The system capacity per macro cell area is defined as the maximum offered load that can be tolerated for a certain minimum 5 % outage user throughput (e.g. 4 Mbps). This definition is used for comparing the relative capacity gains with DC as compared to the case without DC. For users configured with DC, another important performance measures are the buffering time in the SeNB and the probability of SeNB buffer being empty. The buffering time in the SeNB is defined as the time that elapses between the instant a data bit reaches the SeNB buffer and the time in which the same bit is first transmitted to the UE.

The system-level simulator has been extensively tested and verified by reproducing various published performance results in the open literature [23, 29, 30]. As we are able to reproduce such results, it gives confidence that the simulator is reliable. In order to ensure statistical reliable results for the end-user throughput, simulations are run for a time duration corresponding to at least 3000 completed calls. This is sufficient to have a reasonable confidence level for both the 5 and 50 % user throughput performance.

5.2 Analysis of flow control

We first analyze how to tune the flow control parameter θ_s . Figure 5 shows the empirical cumulative distribution function (cdf) of the user throughput with different SeNB target buffering time settings. The results are obtained with 20 ms X2 latency, 5 ms flow control periodicity, and 30 Mbps offered load, which corresponds to medium load (~ 40 % PRB utilization on the average). With low value of θ_s , the SeNB only requests a small amount of data from the MeNB at each flow control period, resulting in a higher probability of the SeNB buffer running empty. Thus the throughput gain with DC is compromised. Increasing the

Table 1 Summary of default simulation assumptions

Parameters	Settings/assumptions
Network layout	7 macro sites (21 macro cells), wrap-around
Macro (UMa) [29] ^a	1 cluster of small cells per macro cell, 4 small cells per cluster
LOS pathloss	$22 \cdot \log_{10}(d) + 28.0 + 20.0 \cdot \log_{10}(f)$, shadow fading std = 4 dB, $10 \text{ m} < d < d'_{BP}$
NLOS pathloss	$40 \cdot \log_{10}(d) + 7.8 - 18 \cdot \log_{10}(h_{BS}) - 18 \cdot \log_{10}(h) + 2 \cdot \log_{10}(f)$, shadow fading std = 4 dB, $d'_{BP} < d < 5000 \text{ m}$
LOS probability	$161.04 - 7.1 \cdot \log_{10}(W) + 7.5 \cdot \log_{10}(h) - (24.37 - 3.7(h_{BS})^2) \cdot \log_{10}(h_{BS}) + (43.42 - 3.1 \cdot \log_{10}(h_{BS})) \cdot (\log_{10}(d) - 3) + 20.0 \cdot \log_{10}(f) - (3.2 \cdot (\log_{10}(11.75 \cdot h_{UT}))^2 - 4.97)$, shadow fading std = 6 dB, $10 \text{ m} < d < 5000 \text{ m}$
Small cell (UMi) [29]	
LOS pathloss	$\min(\frac{18}{d}, 1) \cdot (1 - \exp(-\frac{d}{63})) + \exp(-\frac{d}{63})$
NLOS pathloss	$22 \cdot \log_{10}(d) + 28.0 + 20.0 \cdot \log_{10}(f)$, shadow fading std = 3 dB, $10 \text{ m} < d < d'_{BP}$
LOS probability	Shadow fading std = 3 dB, $d'_{BP} < d < 5000 \text{ m}$
Inter-site distance/cluster radius	Macro cell: 500 m (ISD); small cell: 50 m (cluster radius)
Transmit power	Macro eNB: 46 dBm; small cell: 30 dBm
Bandwidth	$2 \times 10 \text{ MHz}$ @ 2 and 3.5 GHz
Transmission mode	2×2 single-user MIMO with rank adaptation
Receiver type	Interference rejection combining with 2 Rx antennas
Antenna gain	Macro: 17 dBi; small cell: 5 dBi; UE: 0 dBi
Antenna pattern	Macro: 3D antenna, 12° down-tilting, crossX-polarized
Bursty traffic model	Small cell and UE: omni directional antenna, crossX-polarized
Packet scheduling	Poisson arrival with fixed payload size of 4 Mbits per UE
UE cell association	Hotspot UE distribution: 1/3 of UEs dropped within the macro cell coverage area 2/3 of UEs dropped within the small cell coverage area (without RE) Cross-carrier proportional fair Without DC: based on A3/A4 RSRQ With DC: PCell based on A3 RSRP, SCell based on A4 RSRQ with threshold = -16 dB
Modulation and coding	QPSK (1/5 to 3/4), 16QAM (2/5 to 5/6), 64QAM (3/5 to 9/10)
BLER target	10 % for the first transmission

Table 1 continued

Parameters	Settings/assumptions
HARQ modeling	Ideal chase combining with max 4 transmissions
Scheduled throughput exchange periodicity	Every 50 ms

^a f is the centre frequency (Hz), h_{BS} and h_{UT} are the actual antenna heights at the BS and the UT (m), h'_{BS} and h'_{UT} are the effective antenna heights at the BS and the UT (m), d is the distance (m), d_{BP} is the break point distance $d_{BP} = 4 \cdot h'_{BS} \cdot h'_{UT} \cdot f / c$ where c is the propagation velocity in free space, W is the street width (m), and h is the average building height (m)

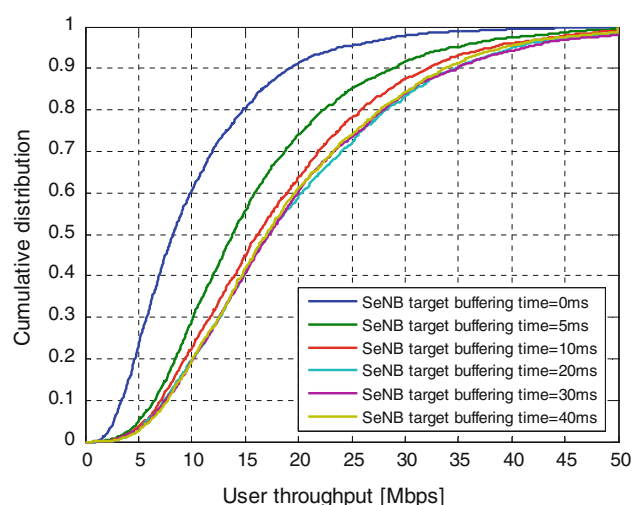


Fig. 5 User throughput with DC under different SeNB target buffering time settings (θ_s), X2 latency = 20 ms, flow control periodicity = 5 ms, offered load = 30 Mbps

Table 2 Probability of SeNB buffer being empty, X2 latency = 20 ms, offered load = 30 Mbps

Target buffering time θ_s	0 ms	5 ms	10 ms	20 ms	30 ms	40 ms
Probability of SeNB buffer being empty	58 %	28 %	14 %	5 %	4 %	3 %

value of θ_s reduces the probability of the SeNB buffer being empty, thus UEs configured with DC have more chance to have simultaneous data reception from the two layers. Generally, the user throughput performance improves with the increase of θ_s , but after a certain point (between 10 and 20 ms in the considered scenario) it remains on a steady level. Table 2 summarizes the probability of SeNB buffer being empty. The probability of the SeNB buffer being empty drops from 58 to 5 % when increasing the value of θ_s from 0 to 20 ms. Further increasing the value of θ_s does not significantly improve the user throughput performance as the probability of SeNB buffer being empty is already very low.

Figure 6 shows the empirical cdf of the buffering time in the SeNB for different values of θ_s . As expected the buffering time increases with the increase of θ_s . That is because the average SeNB buffer size increases with higher values of θ_s , resulting in a longer buffering time. From Figs. 5 and 6, it is observed that with DC there is a trade-off between improving user throughput and reducing SeNB buffering time, which can be balanced by proper configuration of θ_s . With the target of maximizing the throughput while keeping the buffering time at an acceptable level, the optimal setting of θ_s for 20 ms X2 latency and 5 ms flow control periodicity is found to be approximately 20 ms.

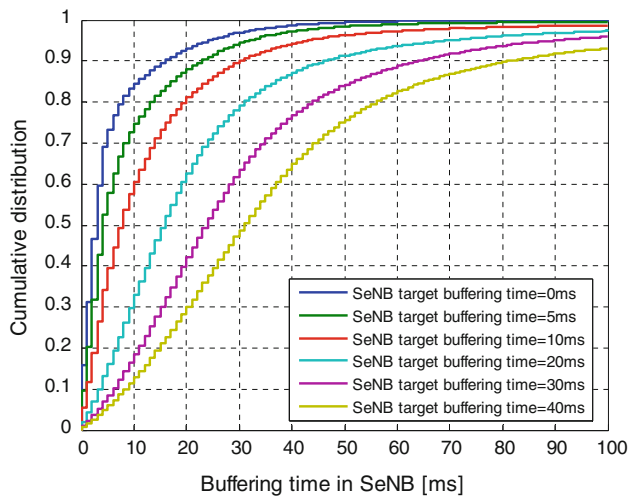


Fig. 6 Buffering time in the SeNB with DC under different SeNB target buffering time settings (θ_s), X2 latency = 20 ms, flow control periodicity = 5 ms, offered load = 30 Mbps

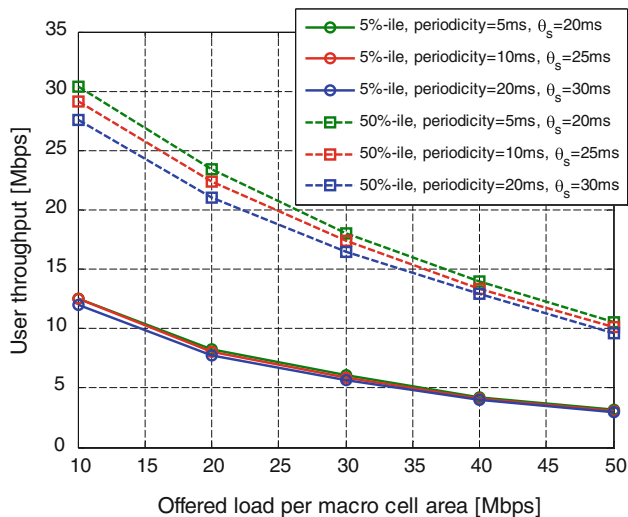


Fig. 7 5 and 50 % user throughput with different flow control periodicity, X2 latency = 20 ms

The impact of the flow control periodicity on the throughput performance with DC is shown in Fig. 7, assuming 20 ms X2 latency. It is observed that for a fixed X2 latency, the value of θ_s has to be increased accordingly with the increase of flow control periodicity (e.g., the setting of θ_s is increased from 20 to 30 ms when the flow control periodicity is increased from 5 to 20 ms). With short flow control periodicity, the MeNB gets frequent status updates (e.g., buffer status and user throughput information) from the SeNB, therefore only the right amount of data is needed at each forward. However with long flow control periodicity, the MeNB gets less frequent status updates from the SeNB and therefore more data has to be forwarded in order to compensate for the fast

Table 3 Flow control parameter settings with traditional backhaul

Parameters	Settings		
X2 latency (ms)	5	20	50
SeNB target buffering time θ_s (ms)	15	20	20
Flow control periodicity (ms)	5	5	5

variations of the instantaneous user throughput in the SeNB. With proper flow control parameter settings, it is shown in Fig. 7 that the 5 % user throughput is not sensitive to the flow control periodicity, while the 50 % user throughput is a bit better with shorter periodicity. In the following simulations, 5 ms flow control periodicity is used as a default setting.

Similarly, simulations with other X2 latencies have been run and an analogous trade-off between user throughput and buffering time in the SeNB was observed. The recommended settings of θ_s and flow control periodicity for different X2 latencies are listed in Table 3. Those recommended values are used as default settings in the following simulations. Again the optimization criterion was to find a proper balance between maximizing the user throughput and keeping the buffering time in the SeNB at an acceptable level. From extensive simulations it is found that the optimal setting of θ_s depends on the X2 latency (Δ) and the flow control periodicity (ρ). An approximate expression for the setting of θ_s is found to be:

$$\theta_s \cong \min \left\{ \frac{\Delta + 40}{3}, 20 \right\} + 5 \log_2 \left(\frac{\rho}{5} \right) \quad (13)$$

As a rule of thumb, higher value of θ_s has to be used with either large X2 latency or long flow control periodicity in order to compensate for the fast variations of the user throughput in the SeNB. It is also worth mentioning that only best effort traffic is simulated in our study as DC in the form of bearer split is mainly targeted for traffic types with higher data rate but loose delay requirements. For traffic types with tight delay but low data rate requirements such as voice, DC is not an appropriate technique to apply.

Figure 8 shows the empirical cdf of the buffering time in the SeNB under different X2 latencies and traffic loads. Low load (10 Mbps offered load) and high load (50 Mbps offered load) correspond to an average eNB PRB utilization of approximately 10 and 80 %, respectively. It is generally observed that for a given X2 latency the buffering time increases as the load increases, but in any case (even at high load) the median value of the buffering time is well kept around the target buffering time θ_s . It is also observed that the buffering time increases as the X2 latency increases. For example, the buffering time with 50 ms X2 latency is larger as compared to the case with 20 ms X2 latency, even though the target buffering time θ_s is set to be

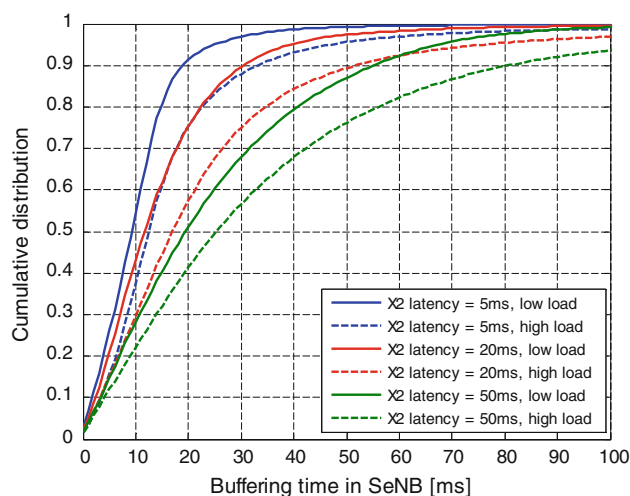


Fig. 8 Buffering time in SeNB with DC under different X2 latencies and traffic loads

20 ms in both cases. That is because with larger X2 latency, the flow control mechanism cannot adapt fast enough to the variations of the user channel quality and load conditions.

5.3 Performance gain of DC

We next evaluate the user throughput gain of DC over traditional backhairs with the flow control parameter settings in Table 3. For cases without DC, only the performance with optimal RE offset at each offered load is plotted as a reference. Figures 9 and 10 show the 5 and 50 % user throughput with and without DC. The performance with inter-site CA and ideal fiber-based fronthaul connection is also plotted as an upper bound of the performance. It is shown that both the 5 and 50 % user throughput with DC are significantly higher than without DC. The gain mechanism with DC is multifold. Firstly, users with DC benefit from higher transmission bandwidth by accessing the resources from the two layers. Quite obviously this bandwidth gain is higher at low load, i.e. when the probability of having a single user accessing all the available radio resources in both the macro and the small cells is higher. Moreover, because users configured with DC can simultaneously be allocated resources in the macro and small cell layers, the system can benefit from increased multi-user scheduling diversity order and faster inter-layer load balancing, thus achieving a better utilization of the radio resources across multiple layers. This is better exploited by users experiencing lower data rates, as cross-carrier PF packet scheduling aims at maximizing the sum of $\log(R_i)$ over all users. That is also why it is observed that the gain of DC at 5 % user throughput is higher as compared to the 50 % user throughput. It is worth mentioning that the gain mechanism with DC is most

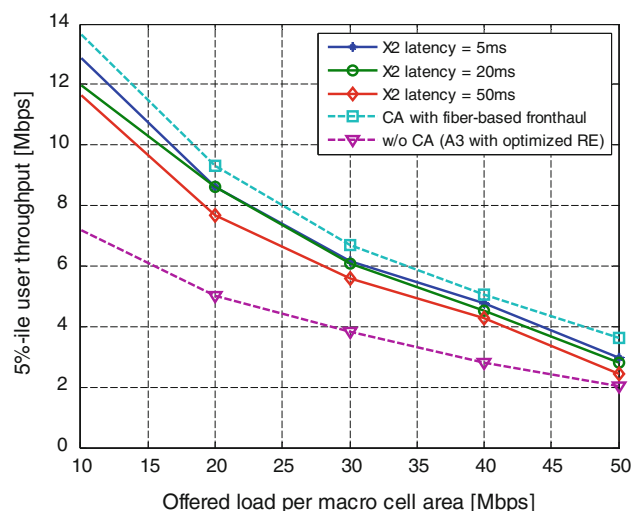


Fig. 9 5 % user throughput with/without DC under different backhaul configurations

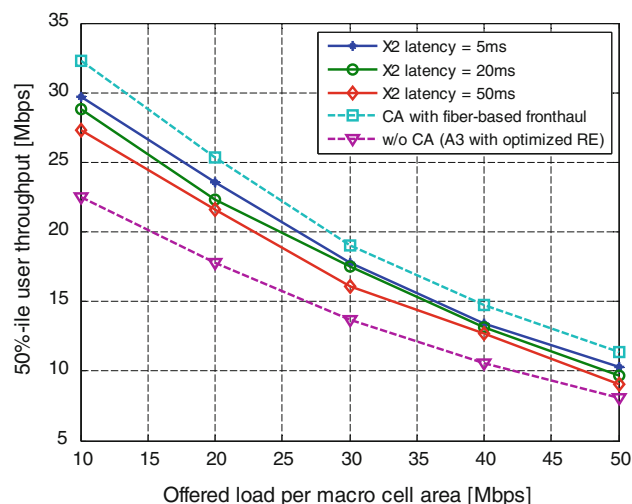


Fig. 10 Median user throughput with/without DC under different backhaul configurations

dominant at low to medium offered load as the throughput gain decreases with the increase of the load.

In general, with the proposed flow control scheme and the recommended parameter settings, the 5 and 50 % user throughput with DC over X2-type backhaul connections are relatively close to the performance with inter-site CA and ideal fiber-based fronthaul connections. The user throughput performance decreases a bit as the latency over X2 increases, but in any case it is significantly better than the performance without DC. From Figs. 7, 9 and 10, it is fair to conclude that the proposed flow control algorithm is robust to adapt to various backhaul configurations and traffic conditions. For a target 5 % outage throughput of 4 Mbps, the maximum tolerable offered load increases from 28 (without DC) to 44 Mbps and 47 Mbps for the cases

with DC and inter-site CA, respectively. This corresponds to a capacity gain of about 60 %. With efficient flow control, DC over X2-type backhaul connections achieves 80 % of the gain available with inter-site CA and ideal fiber-based fronthaul connections.

6 Conclusions

In DC scenarios where the macro and small cells are interconnected with traditional backhaul connections, we have proposed an effective flow control algorithm to forward the data from the MeNB to the SeNB. Besides, general guidelines for the RRM functionalities that most significantly impact the performance of DC, namely UE cell association and packet scheduling, have also been provided. Simulation results show that with DC, there is a trade-off between user throughput and SeNB buffering latency. With the proposed flow control algorithm, such trade-off can be properly balanced by configuring the target buffering time in the SeNB and the flow control periodicity. As the performance of DC under different configurations of X2 latency, flow control periodicity, and traffic load is relatively close to the performance with inter-site CA and fiber-based fronthaul connections, it suggests that the proposed flow control algorithm is generally robust and able to adapt to different conditions. It is shown that the performance of DC over traditional backhaul can achieve 80 % of the gain available with inter-site CA assuming fiber-based fronthaul connections. Specifically, 60 % capacity gain for a target 5 % outage throughput of 4 Mbps is obtained with bursty traffic as compared to the case without DC. The gain with DC comes from larger transmission bandwidth by accessing the two cells as well as increased multi-user diversity and faster inter-eNB load balancing.

Acknowledgments The authors would like to express their gratitude to Jens Steiner and many other colleagues from Nokia Networks and Wireless Communication Networks Section at Aalborg University, Denmark, for valuable discussions and their assistance in the development of the system-level simulator.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Damnjanovic, A., et al. (2011). A survey on 3GPP heterogeneous networks. *IEEE Wireless Communications Magazine*, 18(3), 10–21.
2. Andrews, J. G. (2013). Seven ways that HetNets are a cellular paradigm shift. *IEEE Communications Magazine*, 51(3), 136–144.
3. Ghosh, A., et al. (2012). Heterogeneous cellular networks: From theory to practice. *IEEE Communications Magazine*, 50(6), 54–64.
4. Kishiyama, Y., et al. (2013). Future steps on LTE-A: Evolution toward integration of local area and wide area system. *IEEE Wireless Communications Magazine*, 20(1), 12–18.
5. Nakamura, T., et al. (2013). Trends in small cell enhancements in LTE advanced. *IEEE Communications Magazine*, 51(2), 99–105.
6. Soret, B., Wang, H., Pedersen, K. I., & Rosa, C. (2013). Multicell cooperation for LTE-advanced heterogeneous network scenarios. *IEEE Wireless Communications Magazine*, 20(1), 27–34.
7. 3GPP Technical Report 36.842. Small cell enhancements for E-UTRA and E-UTRAN—Higher Layer Aspects, V1.0.0 (2013–11). Available at www.3gpp.org.
8. Zakrzewska, A., et al. (2013). Dual connectivity in LTE HetNets with split control- and user-plane. In *IEEE proceedings of Globecom workshops, December 2013*.
9. Ishii, H., Kishiyama, Y., & Takahashi, H. (2012). A novel architecture for LTE-B :C-plane/U-plane split and phantom cell concept. In *IEEE proceedings of Globecom Workshops, December 2012*.
10. Mukherjee, A. (2014). Macro-small cell grouping in dual connectivity LTE-B networks with non-ideal backhaul. In *IEEE proceedings of ICC, June 2014*.
11. Wang, H., Rosa, C., & Pedersen, K. I. (2013). Dedicated carrier deployment in heterogeneous networks with inter-site carrier aggregation. In *IEEE proceedings of WCNC, April 2013*.
12. Wang, H., Rosa, C., & Pedersen, K. I. (2013). Analysis of carrier deployment strategies for LTE—A HetNets with multicell cooperation. In *IEEE proceedings of VTC, September 2013*.
13. Pedersen, K. I., Barbera, S., Michaelsen, P.-H., & Rosa, C. (2013). Mobility enhancements for LTE-advanced multilayer networks with inter-site carrier aggregation. *IEEE Communications Magazine*, 51(5), 64–71.
14. Barbera, S., Pedersen, K. I., Michaelsen, P.-H., & Rosa, C. (2013). Mobility analysis for inter-site carrier aggregation in LTE heterogeneous networks. In *IEEE proceedings of VTC, September 2013*.
15. Prasad, A., & Maeder, A. (2014). Energy saving enhancement for LTE-advanced heterogeneous networks with dual connectivity. In *IEEE proceedings of VTC, September 2014*.
16. Tukmanov, A., et al. (2013). On the impact of network geometric models on multicell cooperative communication systems. *IEEE Wireless Communications Magazine*, 20(1), 75–81.
17. Andrews, J., et al. (2010). A primer on spatial modeling and analysis in wireless networks. *IEEE Communications Magazine*, 48(11), 156–163.
18. Pedersen, K. I., et al. (2011). Carrier aggregation for LTE-advanced: Functionality and performance aspects. *IEEE Communications Magazine*, 49(6), 89–95.
19. Zukang, S., et al. (2012). Overview of 3GPP LTE-advanced carrier aggregation for 4G wireless communications. *IEEE Communications Magazine*, 50(2), 122–130.
20. Hytonen, V., et al. (2012). Introduction of multiflow for HSDPA. In *IEEE proceedings of new technologies, mobility and security (NTMS), May 2012*.
21. Yaver, A., et al. (2012). On the joint usage of MIMO and multiflow in evolved HSPA networks. In *IEEE proceedings of international conference on communications (ICC), June 2012*.
22. Dimitri, B., & Robert, G. (1992). *Data networks* (2nd ed.). Englewood Cliffs: Prentice Hall.
23. 3GPP Technical Specification 36.214. Evolved universal terrestrial radio access (E-UTRA); physical layer measurements, V11.1.0 (2012–12). Available at www.3gpp.org.

24. Peng, M., et al. (2013). Self-configuration and self-optimization in LTE-advanced heterogeneous networks. *IEEE Communications Magazine*, 51(5), 36–45.
25. 3GPP Technical Specification 36.331. Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification, V10.5.0 (2012-03). Available at www.3gpp.org.
26. Holtzman, J. M. (2001). Asymptotic analysis of proportional fair. In *Proceedings of IEEE PIMRC, September 2001* (pp. F33–F37).
27. Wang, Y., Pedersen, K. I., Sørensen, T. B., & Mogensen, P. E. (2010). Carrier load balancing and packet scheduling for multi-carrier systems. *IEEE Transactions on Wireless Communications*, 9(5), 1780–1789.
28. Wang, Y., Pedersen, K. I., & Sørensen, T. B. (2011). Utility maximization in LTE-advanced systems with carrier aggregation. In *Vehicular technology conference (VTC Spring)*, May 2011.
29. 3GPP Technical Report 36.872. Small cell enhancements for E-UTRA and E-UTRAN—Physical layer aspects, V12.1.0 (2013-12). Available at www.3gpp.org.
30. ITU-R: Guidelines for evaluation of radio interface technologies for IMT-advanced. Report ITU-R M.2135-1, December 2009. Available at www.itu.int.
31. Algans, A., Pedersen, K. I., & Mogensen, P. E. (2002). Experimental analysis of the joint statistical properties of azimuth spread, delay spread, and shadow fading. *IEEE Journal on Selected Areas in Communications (JSAC)*, 20(3), 523–531.
32. Pedersen, K. I., et al. (2009). An overview of downlink radio resource management for UTRAN long-term evolution. *IEEE Communications Magazine*, 47(7), 86–93.
33. Lampinen, M., et al. (2012). System-level modeling and evaluation of interference suppression receivers in LTE system. In *IEEE proceedings vehicular technology conference (VTC-Spring)*, May 2012.
34. Brueninghaus, K., et al. (2005). Link performance models for system level simulations of broadband radio access systems. In *IEEE 16th international symposium on personal, indoor and mobile radio communications (PIMRC)*, September 2005 (Vol. 4, pp. 2306–2311).



Hua Wang received his Bachelor's degree from Zhejiang University, China, in 2003, and his M.Sc. and Ph.D. degrees from Technical University of Denmark in 2005 and 2009, respectively. From 2005 to 2006 he worked as a radio engineer at Danske Telecom, Copenhagen. From 2008 to 2009 he was a visiting scholar at the University of Texas at Austin. From 2009 he joined Aalborg University as a postdoc. He is currently an Associate Professor at Aalborg

University in collaboration with Nokia Networks, Denmark. His main

research interests are in the area of radio resource management and performance analysis for small cell enhancements in heterogeneous networks.



Claudio Rosa received his M.Sc. E.E. and Ph.D. degrees in 2000 and 2005, respectively, from Aalborg University. In 2003 he also received an M.Sc. E.E. degree in telecommunication engineering from Politecnico di Milano, Italy. He is currently with Nokia in Aalborg, Denmark, where he works as a senior wireless network specialist. At present his research activity is mainly focused on 3GPP standardization of LTE Advanced, more specifically on small cell enhancements and unlicensed spectrum opportunities for cellular deployments.



Klaus I. Pedersen received the M.Sc. degree in electrical engineering and the Ph.D. degree from Aalborg University, Aalborg, Denmark, in 1996 and 2000, respectively. He is currently a Senior Wireless Network Specialist at Nokia Networks, Aalborg, and a Part-Time Professor of radio access technology with Aalborg University. He is the author/coauthor of more than 100 peer-reviewed publications on a wide range of topics, as

well as an inventor on several patents. His current work is related to 5G air interface design, including radio resource management aspects, and Long-Term Evolution and its future evolution.